

Psychological Monographs

General and Applied

o. 376
1954

E. Wandt and L. M. Ostreicher

Validity of Samples of Classroom Behavior
for the Measurement of "Social-Emotional
Climate"

By

Edwin Wandt
and Leonard M. Ostreicher

College of The City of New York

Price, \$1.00



Edited by Herbert S. Conrad
Published by The American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology*
with the *Psychological Monographs*

Editor

HERBERT S. CONRAD

Department of Health, Education, and Welfare
Office of Education
Washington 25, D.C.

Managing Editor

LORRAINE BOUTHILET

Consulting Editors

DONALD E. BAKER

FRANK A. BEACH

ROBERT G. BERNHOUTER

WILLIAM A. BROWNELL

HAROLD E. BURTT

JERRY W. CARTER, JR.

CLYDE H. COOKE

JOHN G. DARKEN

JOHN F. DANIELL

EUGENIA HANFMAN

EDNA HEDBERG

HAROLD E. JONES

DONALD W. MACKINNON

LORRIN A. RIGGS

CARL R. ROGERS

SAUL ROSENZWEIG

ROSS STAGNER

PERCIVAL M. SYMONDS

JOSEPH TIFFIN

LEDYARD R. TUCKER

JOSEPH ZUBIN

Manuscripts should be sent to the Editor.

Because of lack of space, the *Psychological Monographs* can print only the original or advanced contribution of the author. *Background and bibliographic materials must, in general, be totally excluded* or kept to an irreducible minimum. Statistical tables should be used to present only the most important of the statistical data or evidence.

The first page of the manuscript should contain the title of the paper, the author's name, and his institutional connection (or his city of residence). Acknowledgments should be kept brief, and appear as a *footnote* on the first page. No table of contents need be included. For other directions or suggestions on the preparation of manuscripts, see: CONRAD, H. S. Preparation of manuscripts for publication as monographs. *J. Psychol.*, 1946, 26, 447-459.

Correspondence concerning business matters (such as author's fees, subscriptions and sales, change of address, etc.) should be addressed to the American Psychological Association, Inc., 1350 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

Copyright, 1946, by THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Psychological Monographs: General and Applied

Validity of Samples of Classroom Behavior for the Measurement of "Social-Emotional Climate"

Edwin Wandt and Leonard M. Ostreicher

Office of Research and Evaluation, Division of Teacher Education, College of the City of New York

STATEMENT OF THE PROBLEM

OBSERVATIONAL techniques have been widely used in education for many years. Educational research workers as well as administrators and supervisors employ a variety of observational methods to obtain information regarding the classroom behaviors of teachers and their classes. The validity of the generalizations made on the basis of these observational data depends, in the last analysis, on the extent to which certain fundamental assumptions are satisfied. The usual assumptions made are that: (a) the presence of the observer does not materially affect the behaviors observed, and (b) the behaviors observed in a teacher's class are representative of those which would have been observed had the observations been made at other times and with other classes.

Since these assumptions usually are not explicit, they are seldom recognized. As a result, generalizations made from such observational data may be of questionable validity. Instances where the assumption of representativeness is made without adequate evidence are legion in day-to-day school situations. It is common for a principal or supervisor who observes a teacher once or twice during a semester to presume that similar behaviors would have been observed had the observations been made on other days, in different activities, or with different groups of children.

The literature contains some suggestions that certain types of teacher performance may be relatively homogeneous, but the evidence is fragmentary and inconclusive. A study by Anderson, Brewer, and Reed (2) revealed a persistence in behavior patterns of two teachers who were studied during two consecutive years with different classes each year. Withall (9, 10) studied the psychological climate of the classroom by means of an analysis of teachers' verbal behavior. Although he did not study variability per se, he concluded that ". . . there appears to be some consistency in the kind of atmosphere the same teacher creates in her classroom over a period of time."

These studies, and many others in the field of education, have employed observations as a means of obtaining samples of behavior. These samples are customarily used as estimates of "typical" behavior. The question, usually ignored, is: *How valid are these estimates?* This problem is not unique to observational data; it is equally applicable to any case in which a sample is used to represent a population. Without a reliable measure of the variability of the data, it is impossible to estimate reliably the precision of the estimate that is made.

The present study was designed to obtain further evidence concerning the variability of teacher and class behaviors. Specifically, answers to the follow-

ing questions were sought:

1. How consistent are the classroom behaviors of teachers and classes over a period of time?
2. Are the behaviors of teachers significantly related to the class (particular

group of pupils) with which the teacher is observed?

3. Are initial observations representative of the average of observations made over a longer period of time?

DESIGN OF THE STUDY AND COLLECTION OF THE DATA

An investigation of the problems outlined in the preceding section must necessarily be concerned with the control of certain variables. These variables may be handled by means of experimental manipulation, or by careful selection of a natural situation in which certain variables are equated simply by the nature of the situation. A report of the Committee on Criteria of Teacher Effectiveness (1) of the American Educational Research Association suggests that research workers in the area of appraisal of teacher effectiveness be alert to the opportunities that such natural situations may provide. It was the search for such a natural situation that led to the selection of a junior high school for this study, although the implications of the study probably transcend that specific level of the educational program.

Setting of the Investigation

A New York City junior high school, in which each class proceeds through the departmental program as an intact group, was selected for the study. The school, situated in a middle-class residential area, serves 1,100 students in grades seven through nine. Within each grade the students are "homogeneously" grouped, primarily on the basis of intelligence.

While there were some apparent differences among the teachers in their philosophical approaches to education, the general demeanor of the school could

be classified as "rather formal." Administrator-teacher relationships seemed to be warm and free from tension.

Design of the Study

In order to investigate variability in teacher behavior it was necessary to employ a design that included repeated observations of teachers with different classes. With this in mind, the investigators selected a high-ability and a low-ability seventh-grade class taught by the same five teachers. The high-ability class was composed of 37 children with a mean IQ of 110; the low-ability class consisted of 27 children with a mean IQ of 80. Neither class could be considered a "problem" class from the standpoint of discipline.

The two classes were observed as they received instruction from the five teachers in art, English, mathematics, music, science, and social studies. One teacher taught both English and social studies. In order to protect the identity of this teacher, the data will be presented as though six teachers had been observed. The experience of the teachers (four female and one male) ranged from approximately four years (for two teachers) to over 30 years for the others. The only basis for selecting these teachers was the fact that they taught the two classes chosen for the study.

Observations for the study proper were scheduled so that each teacher would be observed once with each of

the two classes during each two-week period until seven observations had been obtained for each teacher with each class. Absence of teachers, school holidays, and scheduling demands on the observers necessitated some modifications, but for the most part the routine was carried out as planned. The total study thus consisted of 84 observations of 45-minute duration (six subject matter areas \times two classes \times seven occasions). These 84 observations occurred during the months of January through April, 1953. In 80 of the sessions independent ratings were simultaneously made by the two authors. Illness of one of the authors necessitated the use of single ratings in the remaining four sessions.

Each class was observed in each subject on the same day of the week, and during the same period, on each of the seven observations. For example, the high-ability class was observed with teacher A during the third period on alternate Wednesdays. Since the low-ability class was necessarily observed with teacher A at a different time, this design might tend to reflect differences due to time as well as those due to differences between the classes. Although the time factor could not be controlled, it was the opinion of the observers that this factor was relatively unimportant.

The Rating Scales

It was desired to use, as observational variables, behaviors that commonly enter into the appraisal of teacher performance and are observable at all grade levels and in all subject matter areas. Consequently, it was decided to employ rating scales relating to the social-emotional climate of the classroom.

The rating scales were devised after examination of numerous teacher observation scales reported in the literature.

Those that had greatest influence on the present instrument were Wrightstone's Pupil-Teacher Rapport Scale (11) and Symond's A Series of Rating Scales for Use in the Class Room (8). In several instances, discrete items on the existing scales were grouped to form a single, more general scale on the new instrument. The completed instrument consisted of the 14 scales reproduced in the appendix.

Six scales were related to class behavior and eight to teacher behavior. This class-teacher dichotomy, however, was in large measure arbitrary since both sets of scales reflected the total social-emotional climate in the classroom. Although the "class" scales were scored separately from those that related directly to the teacher, the two sets of scales were not actually psychologically distinct.

There is no analytical method of pre-determining the optimal number of scale units that should be used in a study such as the present one. It is known that too few units may result in overly "coarse" ratings. Too many scale units, on the other hand, may demand of the observer a degree of discrimination that is not in accord with his ability to perceive such fine differences. Symonds (6) believes that seven is the optimal number of scale units for ratings of human traits. He finds that use of more than seven steps produces only a slight increase in reliability. Conklin (3), however, concludes that for a double scale that extends through zero with opposite qualities at the extremes of the scale, nine is the optimal number of scale units. Since some of the scales employed in this study were of this nature, it was decided to use nine-step scales. Each scale was defined by a title and by a description of the behaviors which characterize the

extreme positions. These extreme positions were assigned the values 1 and 9, with the high score indicating behaviors consistent with current educational and mental hygiene theory.

Preliminary Observations

A series of preliminary observations were made to ascertain the reliabilities of the ratings and to give the observers practice in the use of the instrument. For these purposes, a class of eighth-grade pupils in the cooperating school was observed with five of its teachers (other than the teachers observed in the study proper). Three 45-minute observations were made of the class with each of the five teachers, providing a total of 15 observational sessions. Since the two observers independently made simultaneous ratings for each of these sessions, it was possible to secure estimates of the reliabilities of the ratings. These estimates were calculated by computing the correlation between the ratings of the co-observers and applying the Spearman-Brown prophecy formula to obtain an estimate of the reliabilities of the combination of the two ratings. This appeared to be an appropriate procedure since it was planned to use the combined ratings in the analysis of the data for the study proper. Ratings on ten of the scales had estimated reliabilities of greater than .70, while four scales were not as reliably observed. It was decided to employ all of the scales in the conduct of the study proper, to estimate the reliabilities, and to drop from the final analysis of the data any scale not having an estimated reliability of .70 or higher.

Observational Technique

The procedure of rating required that the observers independently assign a

value from one through nine for each of the 14 scales after referring to the descriptions of the end points of the scales. For the purpose of attaining maximum reliability, it would have been desirable for the observers to make notes or ratings during the class sessions. However, the possibility that note taking might prove distracting and perhaps anxiety provoking to the teachers seemed great enough to dissuade the observers from such action. Consequently, the observers delayed their recording of ratings until immediately after leaving the classrooms. It is believed that the facility with which rapport was gained was due in large measure to this procedure and any consequent loss in reliability was negligible.

Rapport with teachers was apparently gained rather quickly. There was little indication that the teachers were "stepping out of character" or "putting on a show" for the benefit of the observers, although that possibility cannot be entirely eliminated in all cases. The students also seemed to recognize early in the investigation that the observers posed no threat to them. When, on occasion, a teacher left the classroom, the students were not deterred by the presence of the observers from "letting off a little steam." Apparently the determinant of whether or not an occasion was propitious for violation of school department regulations was whether the teacher, rather than the observers, was in position to note such behavior. Consequently, such transgressions as the use of socially unacceptable language, "spit-ball" throwing, etc., took place, although rarely, in front of the observers even though the pupils evidently thought it unwise to act in like fashion within range of the teacher's sight or hearing.

ANALYSIS OF THE OBSERVATIONAL DATA

Reliability of the Ratings

The first step in analyzing the data was to estimate the reliability of the combined rating (based on ratings from the two observers) for each scale. This was accomplished by computing the product-

TABLE I
ESTIMATED RELIABILITY COEFFICIENT OF THE
COMBINED RATING FOR EACH SCALE
(*N* = 80)

Scale	Reliability Coefficient
Class Scales	
Evidence of child-child cooperation	.88
Degree of class participation	.85
Class interest in activity	.73
Degree of class freedom*	.65
Evidence of class tension	.80
Class feeling toward teacher	.73
Teacher Scales	
Teacher's use of positive motivational devices	.80
Teacher's use of negative motivational devices	.89
Delegation of responsibility by teacher	.83
Teacher's influence in decision making	.86
Teacher's provision for individual differences*	.65
Teacher's feeling toward class	.82
Evidence of tension of teacher	.74
Evidence of organized planning by teacher*	.43
Composite Scores	
Composite of class scales	.88
Composite of teacher scales	.88
Composite of all scales	.90

* Not used in composite scores or in subsequent analyses.

moment coefficient of correlation between the two individual ratings and applying the Spearman-Brown prophecy formula to the obtained coefficient. The estimated reliabilities of the 14 scales are reported in Table 1. In accordance with a prior decision, the three scales having estimated reliabilities of less than .70 were eliminated from the subsequent analyses.

Combination of the Ratings

Each of the individual scales measured one aspect of the social-emotional climate of the classroom. Although originally it had been intended that the individual scales should be analyzed separately, it seemed that analyses of composite scores relating to teacher behaviors, class behaviors, and over-all climate would serve the purposes of the study equally well. Therefore the individual scales were combined to form three composite scores: (a) composite of all scales: the sum of the eleven scale scores; (b) composite of class scales: the sum of the five scale scores involving observation of the behavior of the class; (c) composite of teacher scales: the sum of the six scale scores involving observation of teacher behavior.

It should be noted that these three scores are not statistically independent, since the composite of class scales and the composite of teacher scales were rather highly correlated ($r=.76$, $N=84$), and since the composite of all scales was a combination of these two. Although it was almost impossible to describe the total climate in a few words, a short description was necessary. To serve this need, the words "harmonious" and "children-centered" were used to describe the climate indicated by high scores on the composite scales. A fuller understanding of the meaning of the scores on these scales can be obtained by examining the descriptions of the end points of the individual scales. The reliabilities of the three composite scores were estimated in the same manner as those of the individual scales. These reliabilities are reported in Table 1.

Since the ratings were obtained from

co-observations, any differences between the means, variabilities, and shapes of the two observers' distributions of ratings for any scale or for the composite scores were necessarily a result of differences between the observers. For the purposes of the study it was desired to eliminate such personal differences, and the ratings of each of the observers (for each of the scales and for each composite score) were therefore converted into normalized standard scores having a mean of 50 and a standard deviation of 10. Ratings of the two observers were subsequently added to give a single value on each scale for each observation. All subsequent analyses of the data were made with these values.

Variability of Social-Emotional Climate on Different Occasions

The data for the 84 observations are

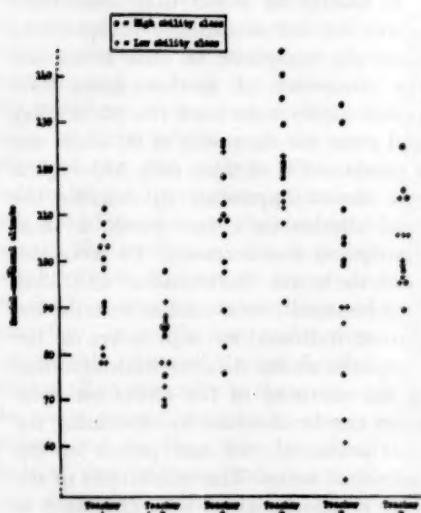


FIG. 1. Ratings on composite of all scales.

(Note: In Fig. 1, 2, and 3 each symbol represents the social-emotional climate on one occasion based on the combined scores from two observers. High scores represent "harmonious" and "children-centered" climates.)

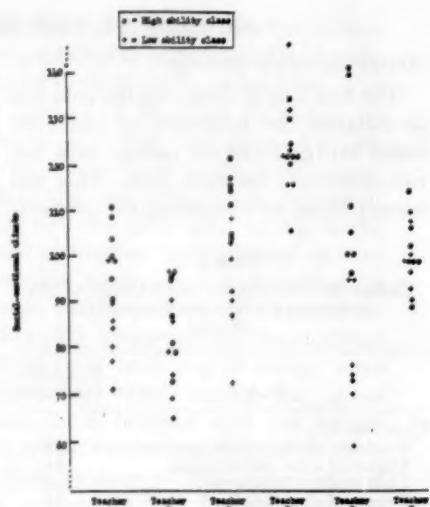


FIG. 2. Ratings on composite of class scales.

shown graphically for the three composite scores in Figures 1, 2, and 3.

From an inspection of the figures, it is apparent that the observed teachers differed in the average climate they maintained. Since it is a truism that teachers

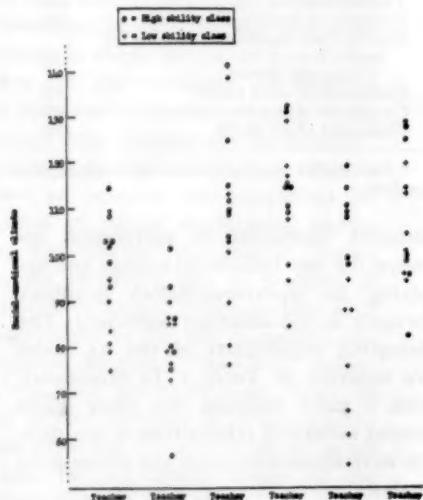


FIG. 3. Ratings on composite of teacher scales.

do differ, this was to be expected. No formal study was made of the differences between the average climates maintained by the teachers since that was not one of the purposes of the study.

The first major question of the present investigation involved the variability of social-emotional climate on different occasions. Although the scores represented in the figures were based on normalized standard scores having no absolute meaning, it was the opinion of the investigators that the range of observed behaviors was quite large. Evidence of an extensive range in an absolute sense came from the fact that with a total possible range of raw scores (before conversion to standard scores) of 22 to 198, the observed raw scores on the composite of all scales ranged from 62 to 154.

Inspection of Figure 1 shows that the climates observed in most of the teachers' classrooms varied considerably during the 14 observations. Similar variation is shown in Figures 2 and 3. The figures also reveal marked differences between the ranges of the observed climates maintained by the various teachers. On the composite of all scales (Figure 1) teacher E had a range of 82 points as contrasted with a range of only 30 points for teacher B. In fact, teacher E had a range approaching in magnitude that exhibited by all the teachers observed (93 points). For teacher E, this extreme range was a function of the differences in behaviors that were observed with the two classes. It may be seen with teachers C and D, however, that the behaviors observed with the only low-ability class covered a very large range.

Comparison of the Climate in the Two Classes

The second major question of the study was concerned with the problem

of whether the social-emotional climate in a given teacher's classroom differed systematically with different classes. To answer this question, the null hypothesis (that there is no difference between the climates observed in the high- and low-ability classes) was tested for each of the teachers. A nonparametric statistic described by Kendall (4) was used to test the hypothesis. This statistic, tau, a measure of rank correlation, can be utilized in various ways; in the present case it provided a test of the null hypothesis between two qualities, one of which was a

TABLE 2
TEST OF NULL HYPOTHESIS FOR COMPOSITE
SCORES OF HIGH- AND LOW-ABILITY
CLASSES

Score	Teacher					
	A	B	C	D	E	F
Composite of All Scales	*	*	*			
Composite of Class Scales	*	*	*			
Composite of Teacher Scales	*				*	

* Null hypothesis rejected ($p \leq .05$). In each of these cases in Tables 2 and 3, the high-ability class was the more "harmonious" and "children-centered."

ranking and the other a true dichotomy. The 14 observations for each teacher (seven with each class) were ranked and the test applied to determine whether the ranks assigned to each of the classes (the dichotomy) were such that they could have occurred by chance. The results of these tests for the composite scores, given in Table 2, show that there was a systematic difference between the climates in the two classes for teachers A, C, and E. The null hypothesis was also tested for the 11 individual scales. These results are reported in Table 3.

One further interesting fact was revealed in the analysis of the data. There was very low agreement between the ranks of the teachers based on the med-

TABLE 3
TEST OF NULL HYPOTHESIS FOR INDIVIDUAL
SCALE SCORES OF HIGH- AND LOW-
ABILITY CLASSES

Scale	Teacher					
	A	B	C	D	E	F
Class Scales						
Evidences of child-child cooperation	•			•		
Degree of class participation				•		
Class interest in activity	•	•	•	•		
Evidence of class tension	•	•	•	•		
Class feeling toward teacher	•	•	•	•		
Teacher Scales						
Teacher's use of positive motivational devices						
Teacher's use of negative motivational devices	•	•	•	•	•	•
Delegation of responsibility by teacher						
Teacher's influence in decision making	•			•		
Teacher's feeling toward class	•	•	•	•		
Evidence of tension of teacher		•	•	•		

* Null hypothesis rejected ($p \leq .05$).

ians of the seven observations with the high-ability class and the ranks based on the observations with the low-ability class. The rank correlation (rho) was .43 for the composite of class scales, .31 for the composite of teacher scales, and .43 for the composite of all scales. None of the foregoing correlations was significant at the .05 level. Thus the rankings of teachers on the basis of classroom climate was quite different for the two classes.

CONCLUSIONS AND IMPLICATIONS

On the basis of the behaviors of the teachers and classes observed during the course of the study, the following conclusions appear to be in order: (a) Social-emotional climate in the classrooms of the observed teachers varied widely from occasion to occasion. (b) Social-emotional climate in the classrooms of three of the

Validity of The Initial Observations

The final question posed in the study concerned the accuracy of initial observations as estimates of typical behavior. To answer this question, the median of the last six observations of each teacher with each class was used as the most suitable over-all measure of typical behavior. The teachers were ranked, for each of the classes, on this median value. This ranking was then compared with the ranking of the teachers on the basis of the initial observations. The statistic rho was employed to measure the extent of relationship between the two sets of rankings. A summary of the correlations is presented in Table 4.

It is evident that the agreement between the ranking based on initial visits and those based on the median of the six subsequent visits is not very close; none of the correlations is significant at the .05 level (a rho of .81 is required for significance at the .05 level with $N = 6$).

TABLE 4
RANK CORRELATION COEFFICIENT (rho) BETWEEN
INITIAL RATING AND MEDIAN OF THE
LAST SIX RATINGS
($N = 6$)

Scale	Class	
	High Ability	Low Ability
Composite of Class Scales	.49	.54
Composite of Teacher Scales	.20	.49
Composite of All Scales	.37	.71

observed teachers varied systematically for the two classes observed. In each case the more "harmonious" and "children-centered" climate was observed in the higher-ability class. (c) Initial observations were unreliable indices of the "typical" climate, even when the class was held constant.

The foregoing conclusions hold whether teacher behavior, class behavior, or a combination of both was used as the estimate of social-emotional climate. It is believed that the implications of this study apply equally well at all levels of teaching, although this has not been demonstrated. It should be noted that the five teachers observed were an accidental sample, having been selected for the sole reason that they happened to teach the two classes chosen for the study.

A recent study by Mitzel and Rabnowitz (5), which utilized Withall's (9) technique for assessing social-emotional climate, reached a similar conclusion with respect to the variability of teachers' verbal behavior. One of their conclusions was that the teachers they observed exhibited marked variations in their verbal behavior on different occasions in addition to differing among each other in typical or average behavior. Further evidence of this variability is found in Withall's article (10). Although he indicated some consistency in the kind of atmosphere created by the same teacher over a period of time, an analysis of his data by the authors revealed marked variability on the part of some teachers.

The results of the present study suggest that if one or two observations are to be used as the basis for making an estimate of a teacher's "typical" behavior, it must first be demonstrated that the behaviors in question are relatively constant from occasion to occasion. If, however, it is ascertained that the behaviors in question do vary considerably, it is possible that a measure of this variability will prove to be of equal or greater

interest than a measure of "typical" behavior.

The systematic differences found in the behaviors of some teachers with different kinds of classes suggest the necessity of considering this factor in studying teacher performance. In many cases, it will be impossible to secure an adequate picture of a teacher's behavior without making repeated observations of that teacher with different groups of students. The implication of these conclusions for research related to the appraisal of teacher performance seems to be self-evident. Studies in this area that take account of the implications of the present investigation will necessarily be time-consuming and expensive, but the gain in validity should make the additional effort profitable.

In respect to administrative appraisal of teacher performance, the implications of this study are equally important. In many school systems a teacher must periodically be rated by supervisors. Lack of time often prevents the securing of a sufficiently large number of observations. In elementary schools, there is the additional problem that in most cases a teacher works with the same group of children for at least one year. Rating a teacher on her behavior with one group of students may give a quite misleading estimate of her over-all ability, and may be, in some respects, a very unfair procedure. Until these considerations are taken into account by educational administrators and supervisors, their ratings, in many instances, may have questionable validity.

SUMMARY

Repeated classroom observations of junior high school teachers and classes were made for the purpose of providing evidence on the following questions: (a) How consistent are the classroom behaviors of teachers and classes over a period of time? (b) Are the behaviors of teachers significantly related to the class (particular group of pupils) with which the teacher is observed? (c) Are initial observations representative of the average of observations made over a longer period of time?

Two seventh-grade classes, one consisting of pupils of high ability and the other consisting of pupils of low ability were observed with five teachers by whom they were taught in common. Four of the teachers taught art, mathematics, music, and science, respectively. The fifth teacher taught two subjects to these classes (English and social studies); in order to protect the identity of this teacher, the data were presented as though six teachers had been observed.

Each class was observed in each subject on seven different 45-minute visitations. A total of 84 observations was made (6 school subjects \times 2 classes \times 7 visits).

The instrument employed for the observations consisted of 14 rating scales.

These scales provided an assessment of the social-emotional climate of the classroom based on the behaviors of the teacher and the class.

Two observers made independent simultaneous ratings on the 14 scales. The ratings of the two observers were later combined in order to yield increased reliability. The individual scales were combined to provide three composite measures of social-emotional climate.

Analysis of these data resulted in the following conclusions: (a) Social-emotional climate in the classrooms of the observed teachers varied widely from occasion to occasion. (b) Social-emotional climate in the classrooms of three of the observed teachers varied systematically with the two classes observed. In each case the more "harmonious" or "child-centered" climate was observed with the higher-ability class. (c) Initial observations were unreliable indices of the "typical" climate, even when the class was held constant.

The implications of these findings for educational research and for administrative use of observations of teachers were discussed.

APPENDIX

RATING SCALES EMPLOYED IN THE STUDY

(Nine-step numerical scales with descriptions of the end points)

*Class Scales**Scale 1: Evidences of Child-Child Cooperation*

(1) No cooperative work; children work as individuals. (9) Almost all pupils engage in cooperative activity; help one another with work; plan together.

Scale 2: Degree of Class Participation

(1) Very few children actively engaged in class activity; participation monopolized by few; little opportunity for full participation. (9) Maximum degree of pupil participation; almost all pupils actively engaged in class activity.

Scale 3: Class Interest in Activity

(1) Vast majority of pupils bored, restless, dislike activity. (9) Vast majority of pupils interested in activity, enthused, eager.

*Scale 4: Degree of Class Freedom**

(1) Pupils remain in designated location; do not speak freely to classmates. (9) Pupils move freely about the room; speak freely to classmates.

Scale 5: Evidence of Class Tension

(1) Vast majority of pupils very tense, many nervous mannerisms, abnormally quiet or irritable. (9) Practically all pupils very relaxed, at ease, natural, "at home."

Scale 6: Class Feeling Toward Teacher

(1) Majority of children appear to dislike teacher; frequent evidence of hostility toward teacher. (9) All children evidently very fond of teacher; react to teacher as friend.

*Teacher Scales**Scale 7: Teacher's Use of Positive Motivational Devices*

(1) Teacher gives no praise or encouragement; no tangible rewards. (9) Teacher gives very many compliments, bestows lavish praise, frequent encouragement.

Scale 8: Teacher's Use of Negative Motivational Devices

(1) Teacher continually threatens or administers some form of punishment; belittles or uses sarcasm. (9) Teacher does not threaten or administer punishment; no veiled threats.

Scale 9: Delegation of Responsibility by Teacher

(1) Teacher does everything for the class; she does all blackboard work, passes supplies, collects papers, keeps records. (9) Teacher delegates great amount of responsibility to children; they write on blackboard, administer materials, keep records.

Scale 10: Teacher's Influence in Decision Making

(1) Teacher makes all decisions. (9) Pupils make decisions on planning and executing activities; Teacher acts as advisor when asked.

* Not used in analysis of data.

Scale 11: Teacher's Provision for Individual Differences*

(1) Teacher expects all children to behave identically; to complete work and comprehend at same time. (9) Teacher gives maximum allowance for individual differences; makes individual assignments.

Scale 12: Teacher's Feeling Toward Class

(1) Teacher shows active dislike toward many pupils; often becomes angry or disgusted with them. (9) Teacher is obviously very fond of the vast majority of the children.

Scale 13: Evidence of Tension of Teacher

(1) Teacher under severe strain, distraught, high strung, tense expression, nervous mannerisms. (9) Teacher very relaxed, at ease, natural; feels at home.

Scale 14: Evidence of Organized Planning by Teacher*

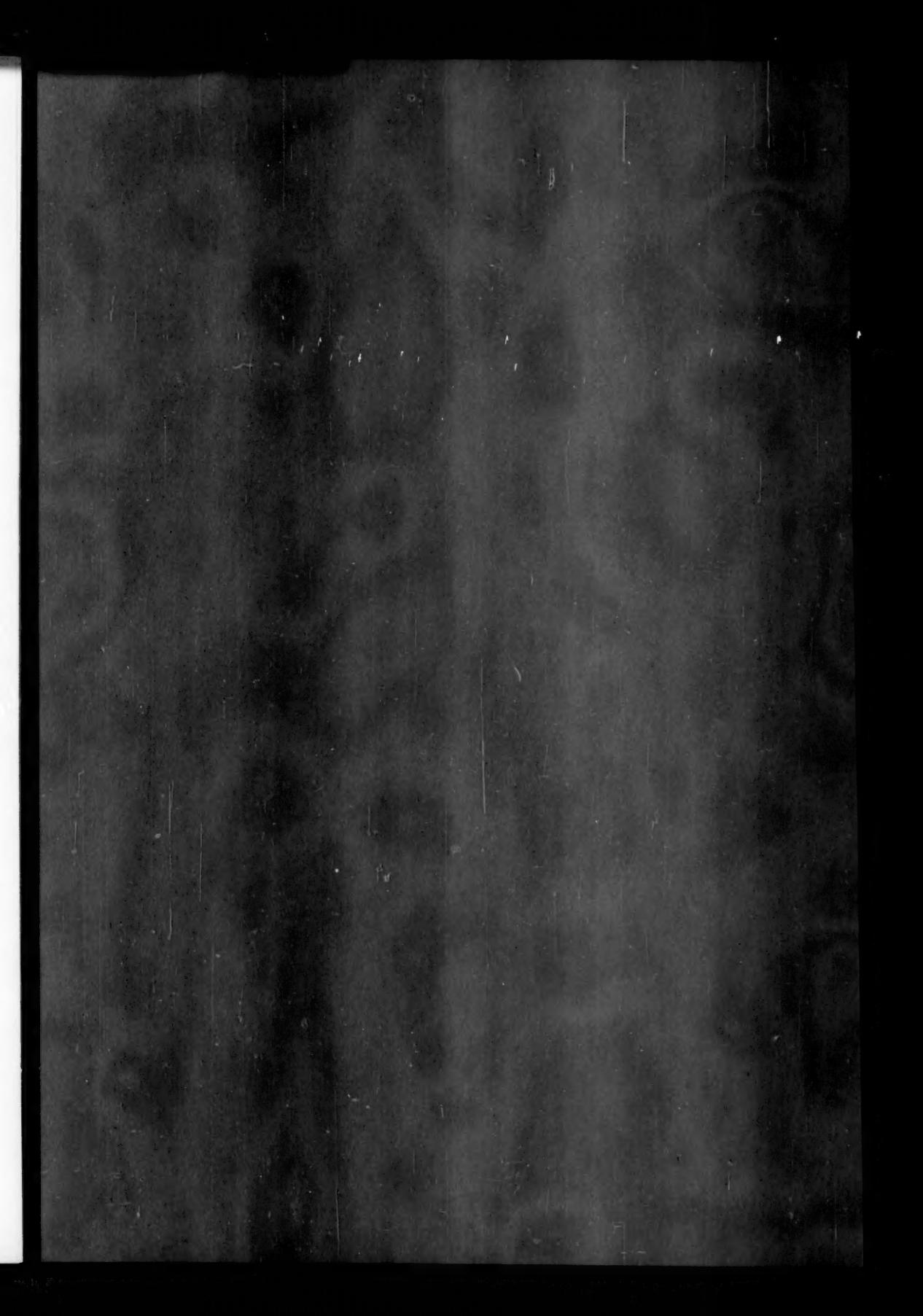
(1) Obvious lack of planning by teacher. (9) Teacher apparently has planned lesson; activities flow smoothly.

* Not used in analysis of data.

REFERENCES

- AMERICAN EDUCATIONAL RESEARCH ASSOCIATION. COMMITTEE ON CRITERIA OF TEACHER EFFECTIVENESS. Report of the Committee on the Criteria of Teacher Effectiveness. *Rev. educ. Res.*, 1952, 22, 238-263.
- ANDERSON, H. H., BREWER, J. E., & REED, M. F. Studies of teachers' classroom personalities, III. Follow-up studies of the effects of domineering and integrative contacts on children's behavior. *Appl. Psychol. Monogr.*, 1946 No. 11.
- CONKLIN, E. S. The scale of values method for studies in genetic psychology. *Univer. Oregon Publ.*, 1929, 2, No. 1.
- KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
- MITZEL, H. E., & RABINOWITZ, W. Reliability of teachers' verbal behavior: a study of Withall's technique for assessing social-emotional climate in the classroom. *Psychol. Monogr.*, 1953, No. 18 (Whole No. 368).
- SYMONDS, P. M. *Diagnosing personality and conduct*. New York: Appleton-Century, 1931.
- SYMONDS, P. M. Reflections on observations of teachers. *J. educ. Res.*, 1950, 43, 688-696.
- SYMONDS, P. M. A series of Rating Scales for use in the Class Room. Unpublished manuscript, Teachers Coll., Columbia Univer.
- WITHALL, J. G. The development of a technique for the measurement of social-emotional climate in classrooms. *J. exp. Educ.*, 1949, 17, 347-361.
- WITHALL, J. G. Assessment of the social-emotional climates experienced by a group of seventh graders as they moved from class to class. *Educ. psychol. Measmt.*, 1952, 12, 440-451.
- WRIGHTSTONE, J. W. Measuring the social climate of a classroom. *J. educ. Res.*, 1951, 44, 341-351.

(Accepted for publication December 2, 1953)



CHICAGO SANTA FE RAILROAD COMPANY, MEMPHIS, TENNESSEE